Multiple Classifier System for Handling Imbalanced and Overlapping Datasets on Multiclass Classification

Dessy Siahaan[1]*, Anwar Fitrianto[2], and Khairil Anwar Notodiputro[3]

[1,3]Statistics and Data Science, Department of Statistics, Faculty of Mathematics and Natural Sciences, IPB University
Jawa Barat, Indonesia 16680
[1]dessysiahaan@apps.ipb.ac.id; [2]anwarstat@gmail.com; [3]khairil@apps.ipb.ac.id

Abstract - The performance of classification models can suffer when the dataset contains imbalanced and overlapping data. These two conditions are already challenging separately and even more complex if they occur together. In the research, an ensemble method called a Multiple Classifier System was proposed to address these issues by combining K-Nearest Neighbour and Logistic Regression. The Synthetic Minority Oversampling Technique (SMOTE) method was also applied to balance the dataset. The One Versus One (OVO) decomposition technique helped the multiclass classification process. A simulation with 18 scenarios proves that the MCS-SMOTE model can handle these problems by providing good performance. The model's performance is also tested using empirical data on Poverty in West Java in 2021. Empirical data also shows that the proposed method performs well, with an accuracy rate of 80.09%, an F1 score of 0.782, and a G-Mean of 0.242. The areas with the highest poverty rates are Bogor, Bekasi City, Bandung City, Bekasi Regency, and Depok City, located near DKI Jakarta, the capital city. Based on existing predictor variables, poor households in West Java are more likely to occur when they do not have access to credit, the number of household members is more than three, multiple families live in one building, and the head of the household has not graduated from elementary school.

Keywords: Multiple Classifier System (MCS), imbalanced datasets, overlapping datasets, multiclass classification

I. INTRODUCTION

Classification is a technique to predict outcomes by categorizing data according to algorithms that use categorical response variables. There are two types of classification: (1) binary, which involves two class categories, and (2) multiclass, which involves three or more classes. The multiclass classification can be more complicated than the binary one because it involves complex interaction patterns. It can become even more challenging when the data are imbalanced. Some classes have many observations, while others have fewer (Tanha et al., 2020). Additionally, if the data overlaps, they have the same characteristics even though they come from different classes. Hence, the classification process becomes even more complex (Lango & Stefanowski, 2022).

It has been observed that dealing with two types of data problems, namely imbalanced and overlapping, can create difficulties in the classification process. When both conditions are present in the same dataset, the difficulty level increases further (Ishak et al., 2022). While many researchers have worked on resolving these problems in binary classification, there is limited research available for multiclass classification because forming a model becomes more challenging when there are more than two class categories.

Kalid et al. (2020) developed a Multiple Classifier System (MCS) to address the issue of imbalanced and overlapping data in credit card frauds and credit card as default payments datasets. The model was created by combining two single classifiers, C4.5 and Naive Bayes. The ensemble model outperformed other single classifiers. In a separate study, Vuttipittayamongkol et al. (2021) generated 1,010 simulated datasets that combined imbalanced and overlapping data conditions. Random Forest was used for classification, which was considered robust against overfitting because it consisted of a collection of simple trees trained independently. The results concluded that classification errors increased in overlapping data, especially when overlapping and imbalanced conditions occurred in the same dataset.

Meidianingsih and Meganingtyas (2022) utilized One Versus One (OVO) as a technique to break down multiclass issues into binary ones, which could be addressed with a binary classifier. The data utilized for simulation contained the level of imbalance and the number of minority classes. As the number of minority classes in multiclass problems increased, complexity also rose (Fernández et al., 2018).

Rosita et al. (2022) studied various multiclass scenarios with imbalanced conditions. They compared the performance of single and ensemble classifiers and implemented the Synthetic Minority Oversampling Technique (SMOTE) method to solve the imbalance problems. The findings suggest that the ensemble approach, especially when combined with SMOTE, was better suited for handling these issues. In another study, Aldania et al. (2023) performed multiclass classification on simulation data based on the level of overlap and Indonesian Industrial Classification Code (Klasifikasi Baku Lapangan Usaha Indonesia (KBLI)) data. They compared the performance of Catboost and Double Random Forest methods and discovered that Catboost was the better model.

Advanced techniques from previous studies are utilized in the research. These techniques include the OVO decomposition technique for multiclass problems, SMOTE for handling imbalanced problems, and MCS for handling imbalanced and overlapping combinations. The MCS is created using a sequential combination of two single classifiers, with the output of the previous classifier serving as input for the next. K-Nearest Neighbor and Logistic Regression are combined to create the model. The model is applied to simulation and empirical data. To demonstrate the robustness of the model, the researchers compare its performance with a combination of the proposed modeling techniques. The model's success is evaluated based on its ability to outperform other modeling techniques, as well as high evaluation scores. It is anticipated that the proposed method can address classification issues that arise when dealing with imbalanced and overlapping data in multiclass scenarios. By simulating various scenarios, the researchers hope to establish that the method can be applied to real-world situations. In addition, the researchers have employed empirical data to demonstrate that the proposed method produces a satisfactory model.

II. METHODS

Currently, there is no universally accepted measure to determine the degree of imbalance and overlap in a dataset. However, the research combines an approach taken by Meidianingsih and Meganingtyas (2022) for measuring imbalance and Aldania et al. (2023) for measuring overlap. The level of imbalance is divided into three categories: extreme, moderate, and mild, which are determined by the proportion of the minority class compared to the datasets. In addition, the researchers take into account the number of minority classes as a simulation data scenario, as shown in Figure 1, there is a 4-class dataset with 2 minority classes and a moderate proportion of data. Around 10,000 observations are generated, and the number of observations in each class follows the imbalance proportion. Table 1 represents each scenario of imbalanced data with the proportion of observation.

To determine the level of overlap, the researchers use the Euclidean distance between centroids. This distance is categorized into three levels: near, medium, and far. The distances for near, medium, and far are 2, 3, and 4 units, respectively. The closer the centroids are to each other, the higher the level of overlap in the dataset. When the researchers combine the scenario of imbalanced and overlapping data, it produces 18 different datasets, which are described in Table 2.
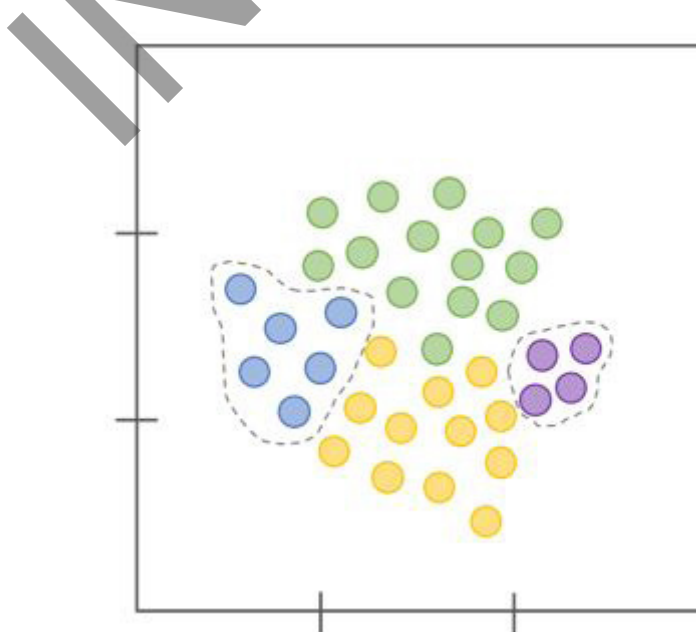

Figure 1 An Example of a Moderate Imbalanced Data Scenario with Two Minority Classes.

Table 1 Simulation Scenarios of the Imbalanced Data

Level of Imbalance	Class A	Class B	Class C	Count of Minority Class(es)
Extreme	9,850	100	50	2
	7,500	2,400	100	1
Moderate	8,900	600	500	2
	7,500	2,000	500	1
Mild	7,800	1,200	1,000	2
	6,500	2,500	1,000	1

Table 2 Simulation Scenarios of the Imbalanced and Overlapping Data

Distance	Minority	Imbalanced
Near	2	Extreme
Near	2	Moderate
Near	2	Mild
Near	1	Extreme
Near	1	Moderate
Near	1	Mild
Medium	2	Extreme
Medium	2	Moderate
Medium	2	Mild
Medium	1	Extreme
Medium	1	Moderate
Medium	1	Mild
Far	2	Extreme
Far	2	Moderate
Far	2	Mild
Far	1	Extreme
Far	1	Moderate
Far	1	Mild

The empirical data used in the research is taken from the National Socio-Economic Survey in West Java Province in 2021. It is secondary data that have been processed into 10 predictor variables and 1 categorical response variable, as shown in Table 3. The response variable describes the level of poverty per household, classified into three categories: not poor, poor, and extremely poor. Poverty is determined based on the inability to fulfill basic needs, measured by the poverty line. In the data, the variable used to categorize each household is the monthly average expenditure per capita. The West Java Poverty Line in 2021 was IDR427,402, meaning households whose average expenditure is below the poverty line are considered poor (Badan Pusat Statistik Kabupaten Pesisir Selatan, 2023). Poverty is a significant social issue not only in Indonesia but across the globe, and it

has become the first goal of sustainable development goals. The World Bank Extreme Poverty Line is $57, equivalent to IDR322,170, with the currency then (Pensasaran Percepatan Penghapusan Kemiskinan Ekstrem, 2022).

The variables used in are chosen based on research conducted by Djamaluddin (2017) on the characteristics of poor households. The regional characteristics are represented by regional type and credit access, while community characteristics are represented by home ownership status and house floor area. Household characteristics are described by the number of household members and families residing in the building. Individual characteristics include the gender, age, and latest education of the household head, as well as the number of hours they work in a week.

There are two common techniques used in classification problems: One Versus All (OVA) and OVO (Esteves, 2020). The OVA method involves splitting the data into binary by selecting one class as the positive class and the remaining classes as the negative class. The OVO method involves dividing the data into binary subclasses with all possible pair combinations based on the number of existing classes. It has been shown that the OVO method is more effective in handling multiclass problems (Galar et al., 2011).

After analyzing the multiclass problem, the next step is to overcome class imbalance. One of the commonly used techniques is resampling, which can be done through under-sampling or over-sampling. Under-sampling involves randomly selecting instances from the majority class to balance it with the minority class. However, this method can lead to a loss of important data. On the other hand, over-sampling involves increasing the data in the minority class, but it can often cause the model to overfit, which means that the model is only good at the training data. The SMOTE method, which involves generating new synthetic data using a distance approach, is proposed to overcome this issue (Chawla et al., 2002).

The classification model starts with the K-Nearest Neighbour (KNN) and Logistic Regression models. Once the model is formed, it is evaluated using three measures: accuracy, F1 score, and recall. However, when dealing with multiclass problems that have class imbalance and overlapping issues, a simple evaluation measure cannot be used. Therefore, instead of using accuracy, F1 score, and recall, the researchers use balanced accuracy, weighted F1 score, and G-Mean.

Both simulation and empirical data undergo the same workflow for model formation and evaluation. The only difference between the two is how the simulation data is generated. Simulation data requires additional steps to generate the data, while empirical data undergoes simple pre-processing. To generate simulation data, the researchers start by generating random numbers for the centroid of class A from a uniform distribution ranging from 0 to 10. For classes B and C, the researchers follow an equilateral triangle pattern to maintain the same distance between them. The centroid generated for class A becomes the first side of the triangle. Then, the researchers determine the desired distance value (near, medium, and far) to regulate the level of overlap.

Next, the researchers generate the centroid for class B by adding the desired distance for the x-axis to class A's centroid and setting the y-axis value to 0. It makes the middle value of class B the other side of the triangle. To generate class C's centroid, the researchers add half of the desired distance for the x-axis and the desired distance multiplied by the square root of 3 divided by 2 for the y-axis to class A's centroid. It makes class C's centroid the vertex of the triangle.

After generating the centroids, the researchers distribute 10,000 samples according to the level of imbalance in each class for each scenario. The researchers then generate numerical variables X_1 and X_2, each from a normal distribution with a mean value modified from previous step and a standard deviation of 1 for each dataset scenario. The researchers also generate categorical variables X_3 and X_4, Variable X_3 has two categories generated from a binomial distribution with probabilities of 0.7 and 0.3. Variable X_4 has three categories generated from a multinomial distribution with probabilities of 0.5, 0.3, and 0.2, respectively.

Once the data has been prepared, it needs to be divided into two groups: training data (80%) for creating the model and testing data (20%) for evaluating it. This division should be done while considering each class. Hence, every class is represented in both training and testing data.

Table 3 Variables on Empirical Data

Code	Variables	Data Type
Y	Poverty Status	Categorical
X_1	Regional Type	Categorical
X_2	Number of Household Members	Categorical
X_3	Gender of Head of Household	Categorical
X_4	Age of Head of Household	Numeric
X_5	Last Education Level of Head of Household	Categorical
X_6	Working Hours by Head of Household per Week	Numeric
X_7	Number of Families Living in a Residential Building	Categorical
X_8	Home Ownership Status	Categorical
X_9	House Floor Area	Numeric
X_{10}	Access to Credit	Categorical

Next, binary subclasses must be formed on each training data using the OVO method. To balance the data, the researchers use the SMOTE technique and build an MCS model using sequential combinations. At the First Level, the researchers classify the data using KNN. The researchers use the results of the First Level's classification as input data at the Second Level. The researchers classify the data using Logistic Regression. Finally, the researchers determine the prediction results using test data.

Next, the researchers resample the data fifty times using an iterative sample selection technique. The researchers use Majority Voting to show predictions for each subclass. Then, the researchers combine the results and use Majority Voting again to determine the final prediction. The final step is to evaluate the model using balanced accuracy, weighted F1 score, and G-Mean.

It is necessary to compare it with simpler models that do not use the MCS combination and SMOTE data balancing to demonstrate the performance of the proposed model. To do this, the researchers follow the same steps as the proposed model, up to the decomposition stage (OVO), for each comparative model. However, the researchers skip the SMOTE data balancing step in the model without data balancing. A single classifier model does not require two modeling stages and Majority Voting. The following are the models used in the research, namely Logistic Regression with SMOTE, KNN with

SMOTE, MCS (KNN-LR) without data balancing, Logistic Regression without data balancing, and KNN without data balancing.

III. RESULTS AND DISCUSSIONS

The proposed model is a sequential combination of KNN and Logistic Regression models and applies the SMOTE method to balance the data between classes. Then, it is compared against five other models to compare the model's performance. The Confusion matrix can be used to measure the model's evaluations. It is a square matrix with rows and columns representing the same category. The rows indicate the amount of predicted data per category, while the columns represent the actual data per category (Brereton, 2021). The Confusion matrix maps the number of observations for each classification result into true positive, true negative, false positive, and false negative. Balanced accuracy, weighted F1 score, and G-Mean are used to evaluate the model's performance.

The data generated for the simulation scenario creation consists of 18 datasets with imbalanced and overlapping data. Each dataset has three class categories, and all stages of model creation are applied to each of them. Figure 2 shows the simulation result in 6 out of 18 scenarios. The evaluation of the proposed model is compared with the evaluation of five other models.
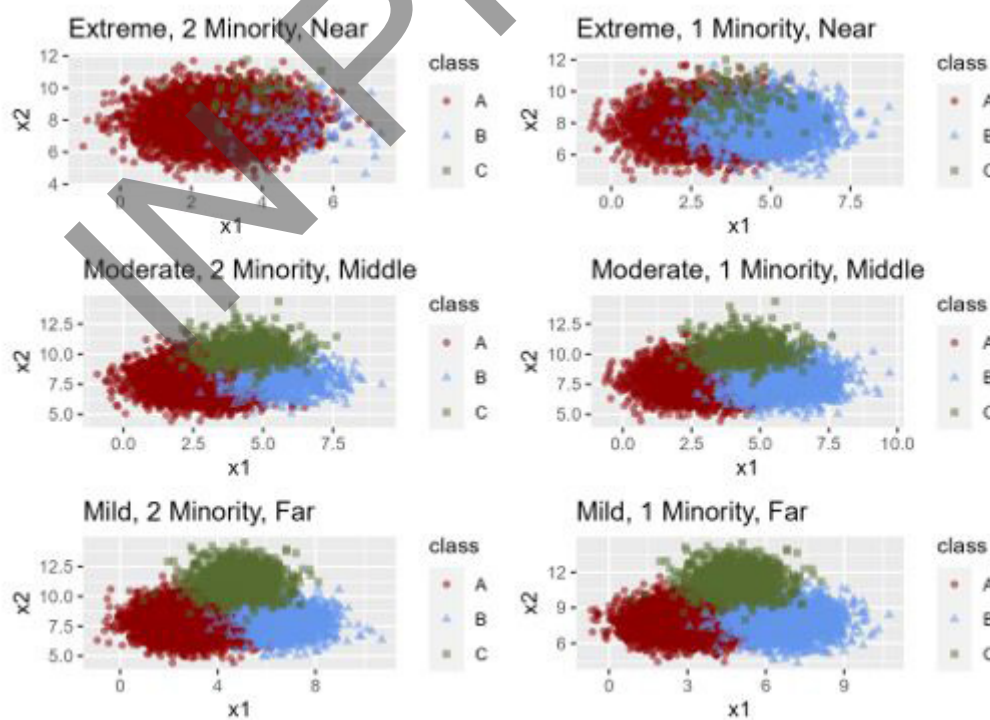

Figure 2 Sample of Simulation Datasets Result

In a multiclass scenario, the accuracy value may not reflect the true performance of a model, as the number of observations in each class may not be balanced. It is suggested to use the balanced accuracy (BA) metric to address this issue. This metric takes into account the model's ability to correctly classify each class by calculating the average sensitivity and specificity values (Equation (1)) across all classes (De Diego et al., 2022). It is achieved by summing the sensitivity and specificity values for each class (i) and dividing by the number of classes (k).

$$BA = \frac{\sum_{i=1}^{k} \text{Sensitivity}_i + \text{Specificity}_i}{k}. \tag{1}$$

According to the experiment shown in Figure 3, the proposed model has demonstrated the best accuracy results among other models, with a percentage value above 75%, except in the Extreme-2 scenario. The low accuracy value in this scenario is due to the high level of data imbalance and the presence of two minority classes, which makes it difficult to classify the data. The KNN-SMOTE model is the second-best model among all the models tested, with its accuracy value being close to the proposed model. The result indicates that SMOTE helps improve model performance in overcoming the problem of imbalanced data, as the model's performance without SMOTE reaches a maximum accuracy of only 55%. The accuracy results improve as the distance scenario increases since a farther distance reduces data overlapping. The reduced overlap of data helps to improve the model's performance.

The F1 score is calculated as the harmonic mean of precision and recall. In a scenario where there are multiple classes, this measure can be expanded using micro-averaging or macro-averaging techniques. However, when there is an imbalance in the data distribution among classes, the evaluation becomes less objective since each class is given the same weight. A weighted F1 score (WF) calculation is used to address this problem. This method assigns appropriate weights to each class (i) based on the amount of data in each class (N), resulting in a more accurate evaluation of the model's performance (Pradana et al., 2022). The formula is in Equation (2).

$$WF = \frac{\sum_{i=1}^{k} \text{F1 Score}_i \times N_i}{N}. \tag{2}$$

It is important to note that not all models are capable of providing an F1 score of 0 because some classes cannot be classified accurately, particularly the true positive. However, the proposed model (MCS-SMOTE) and KNN-SMOTE can classify well, with the proposed model providing the best overall results. The performance of the models has also improved as the distance has increased. In contrast to the accuracy value, the Extreme-2 scenario has shown the best performance in the F1 score. It may be due to the unequal data comparison in the majority class, resulting in a class with a high true positive value. The relatively high true positive value also has an impact on the F1 score produced by the model. Overall, among all the imbalance scenarios, the Extreme-2 scenario has shown the best performance, as shown in Figure 4.
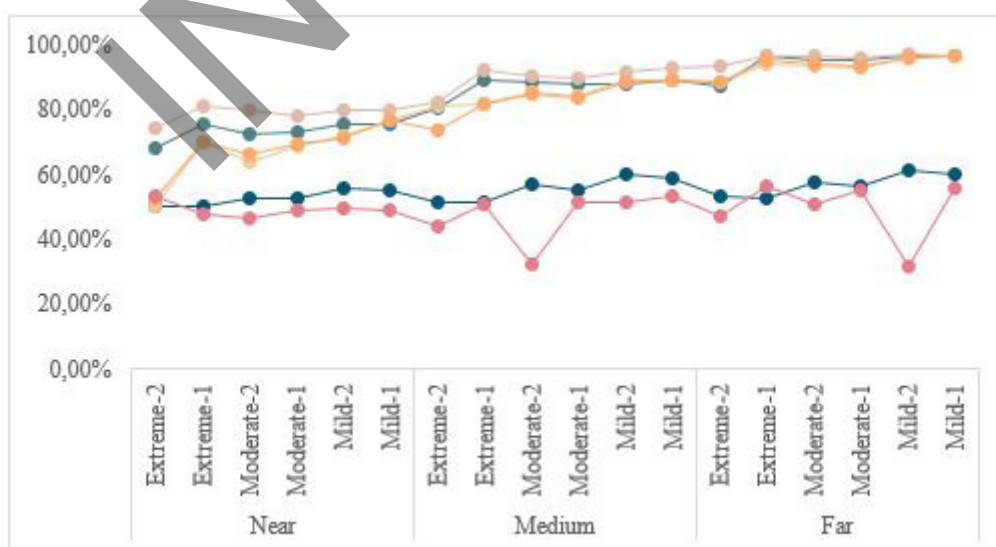

Figure 3 Results of Balanced Accuracy from Six Models on Data Simulation

When dealing with imbalanced data, the generated model may have high accuracy but low sensitivity because a small amount of minority-class data can be examined. Therefore, it is important to pay attention to the sensitivity value. For multiclass data, G-Mean is suggested as the geometric average of recall from each class (Equation 3). G-Mean is capable of measuring the performance produced by the model as it takes into account each acquisition value that represents the classification performance of each class equally (Ongko & Hartono, 2021).

$$\text{G-Mean} = \left(\prod_{i=1}^{k} \text{Recall}_i \right)^{1/k}.$$

(3)

It can be seen from Figure 5 that some models do not provide a G-Mean value as it is calculated using recall, just like the F1 score. However, the proposed model has proven to be the best among all models in all near-distance imbalance scenarios, with a value of over 0.8. The value is considered good performance. It

indicates that the proposed model is effective in dealing with the imbalance problem by accurately classifying minority classes. In the medium-distance scenario, the G-Mean results are better than in the near-distance scenario. However, the Extreme-2 scenario, which has the highest level of overlap and more than one minority class, performs the worst. In the far distance scenario, only the Logistic Regression-SMOTE and KNN models without SMOTE can produce G-Mean values, which significantly improve performance compared to the previous two scenarios. It shows that the level of overlapping has a significant impact on the model formation process.

Through simulation, it has been demonstrated that the proposed MCS KNN-Logistic Regression and SMOTE models perform better than others. These models can handle imbalanced and overlapping data in multiclass classification. However, empirical data are the original data that describe the actual conditions, unlike simulation data, which are generated to meet the desired conditions. Therefore, the proposed model is also tested on empirical data to verify its performance.
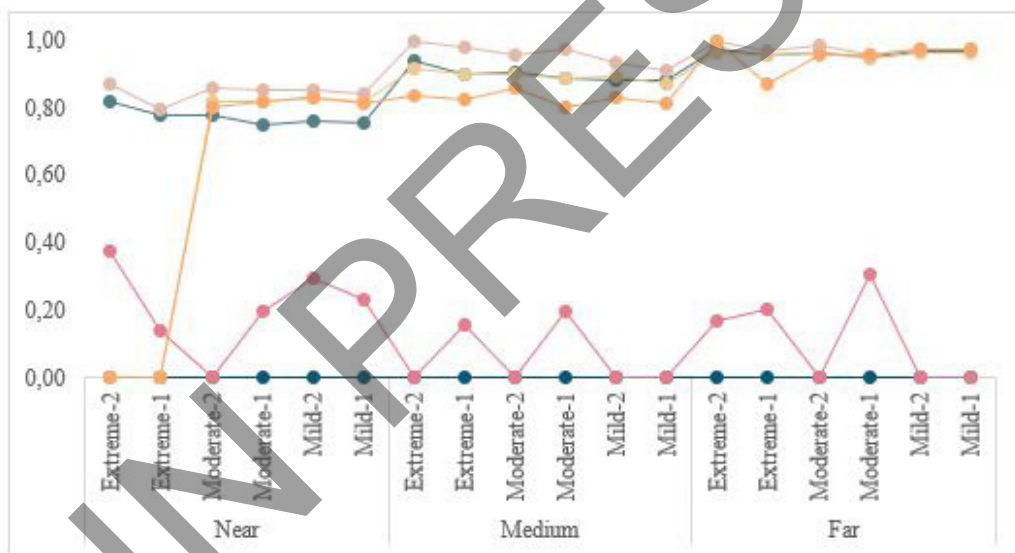

Figure 4 Results of Weighted F1-Score from Six Models on Data Simulation
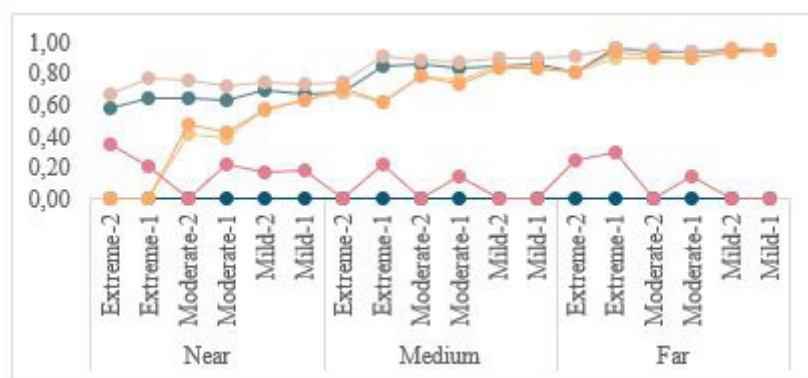

Figure 5 Results of G-Mean from Six Models on Data Simulation

The researchers use data from 25,744 households in West Java. Among them, over 24,000 households are not poor, while fewer than 250 households are extremely poor. According to Figure 6, the level of imbalance observed is similar to the Extreme-2 simulation scenario, where the majority class ratio differs significantly from the minority class, and two minority classes exist, namely poor and extremely poor. Meanwhile, the level of overlapping data is indicated by several numerical variables, as shown in Figure 7.

Next, the process of dividing training data and testing data involves splitting the data proportionally so that each class has the same proportion. The OVO decomposition technique is used to aid the model formation process. In the case of poverty with three classes, the OVO technique divides the data into three binary subclasses: poor and extremely poor, poor and not poor, and extremely poor and not poor.

For each binary subclass, the SMOTE method is applied. In the case of mixed predictor variables (numerical and categorical), the SMOTE-NC method is used to generate new synthetic data for the minority class. The value of $k = 5$ is used to determine the number of nearest neighbours. This process produces balanced data for each binary subclass. Balanced data status is obtained by referring to the majority class in each binary subclass.
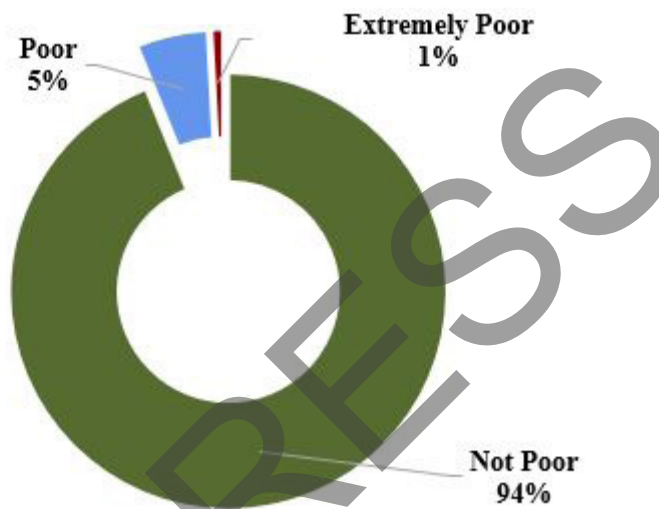

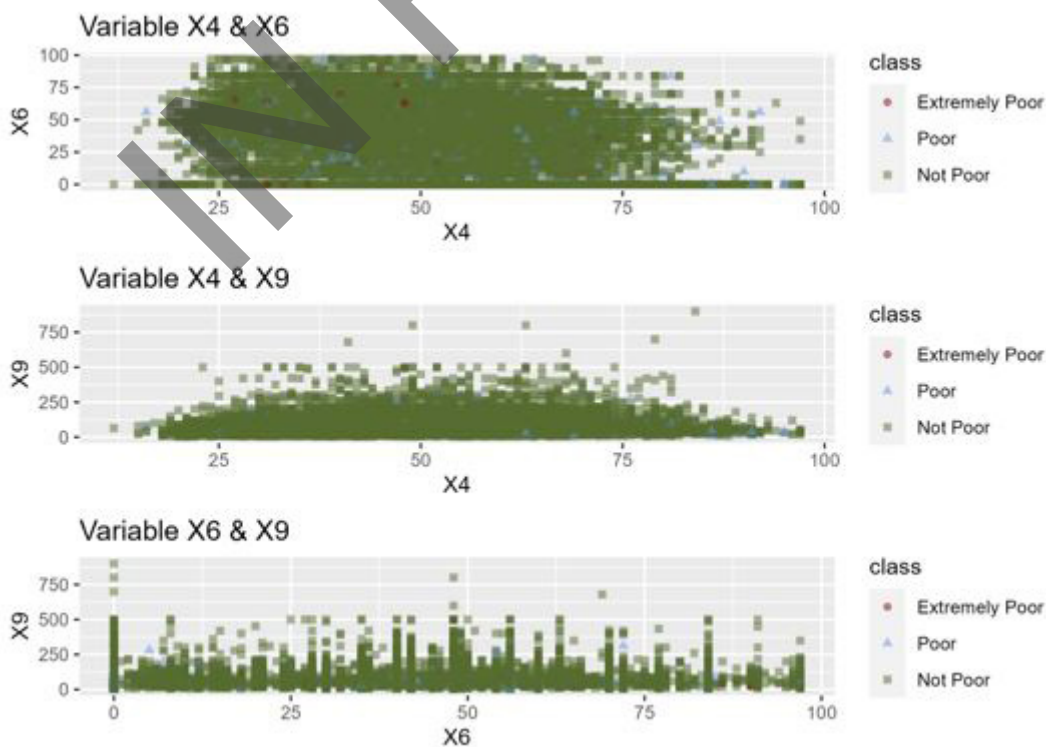Figure 6 Donut Graph of Poverty Status

Figure 7 Overlapping and Imbalanced Data Distribution

After balancing the data in each binary subclass, the first classification stage is KNN modeling. The best (k) value is obtained using cross-validation techniques in each class with the existing function in the caret library in R programming. The KNN model makes predictions for each binary subclass, and the prediction results are saved to replace the response variables in the previous training data, which then become input for the Logistic Regression model.

Logistic Regression modeling is carried out on each binary subclass, and predictions are made on the test data, which is repeated up to 50 times. The final prediction results are determined by majority voting twice. The first stage involves getting prediction results based on the mode value of the 50 repetitions carried out in each subclass. The final prediction result $(\hat{y})$ is a majority vote on test data from each subclass. This method calculates the highest number of classes (mode) produced by each classification model $C_m(x)$ (Sharma et al., 2021). The calculation uses Equation (4).

$$\hat{y} = \text{mode } C_1(x), C_2(x), \cdots, C_m(x) \tag{4}$$

Overall, the proposed model provides the best performance, as presented in Table 4. Almost all comparison models produce F1 and G-Mean scores of 0, indicating poor classification performance. The comparison models fail to classify the minority class correctly, as seen from the G-Mean value. They also fail to classify the entire class accurately. Model evaluation is also carried out on training data to prove that there is no over-fitting in the model formed.

According to the prediction results obtained through the MCS-SMOTE method, the percentage of not poor households is 73.79%, while poor and extremely poor households account for 19.28% and 6.93%, respectively. The poverty level in West Java is quite high, with an even distribution of poverty between urban and rural areas, as shown in Figure 8.
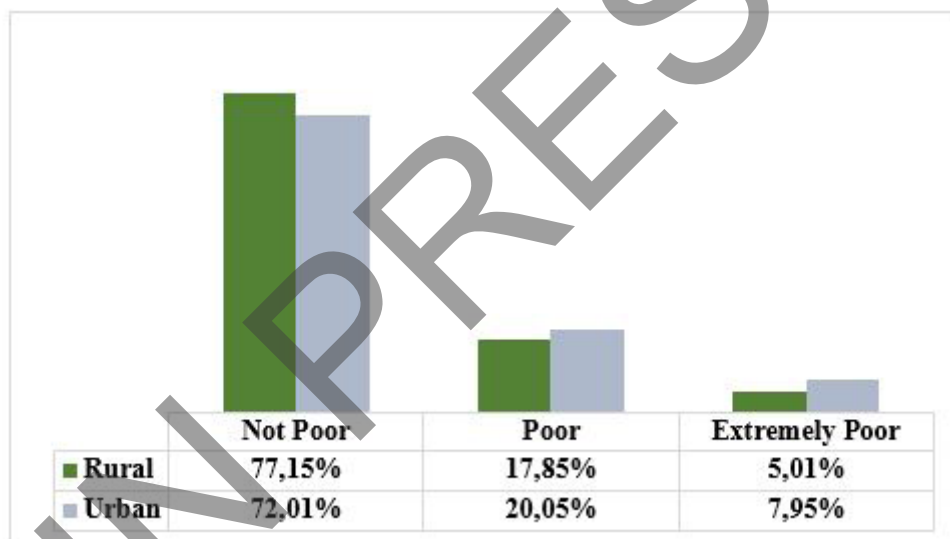

	Not Poor	Poor	Extremely Poor
■ Rural	77,15%	17,85%	5,01%
■ Urban	72,01%	20,05%	7,95%

Figure 8 Prediction of Poverty Status in West Java

Table 4 Comparison of Evaluation Results of Training Data and Test Data for All Models

Classifier	Data Type	Without SMOTE			With SMOTE		
		Accuracy	F1 Score	G-Mean	Accuracy	F1 Score	G-Mean
K-Nearest Neighbour (KNN)	Training	33.90%	0	0	36.27%	0	0
	Testing	36.85%	0	0	34.66%	0	0
Logistic Regression	Training	36.85%	0	0	83.75%	0.817	0.245
	Testing	33.90%	0	0	37.84%	0.053	0.074
Multiple Classifier System (MCS)	Training	36.85%	0	0	80.11%	0.782	0.266
	Testing	36.85%	0	0	80.09%	0.782	0.242

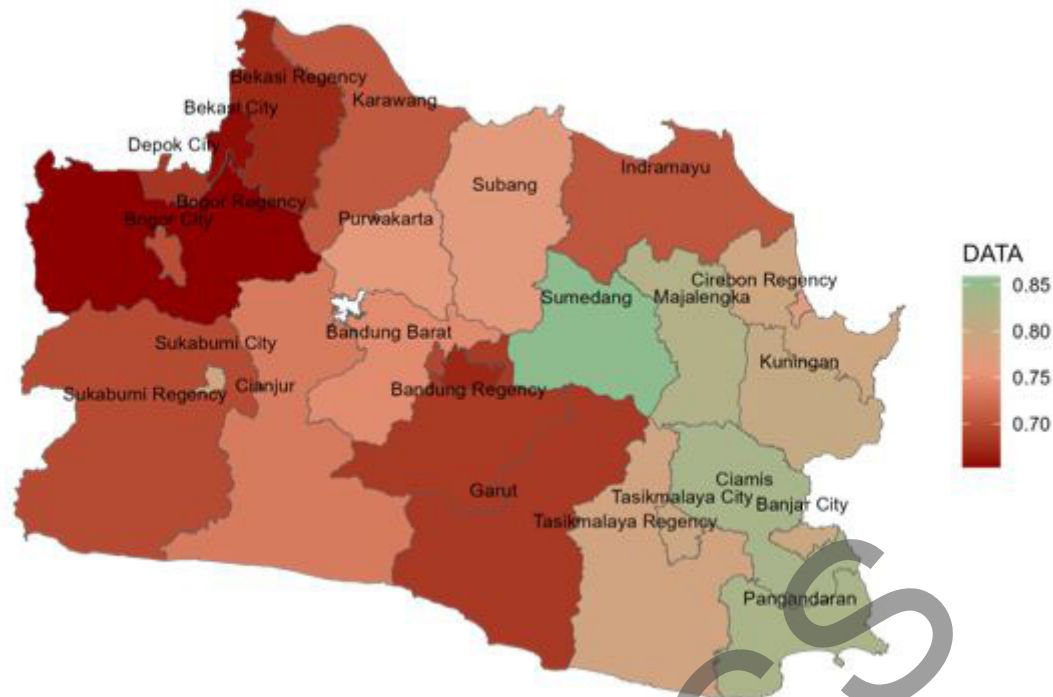Note: Synthetic Minority Oversampling Technique (SMOTE)

Figure 9 Poverty Levels per City/Regency in West Java

Figure 9 displays poverty levels in West Java, with red indicating poorer areas and green indicating more prosperous areas. The highest poverty rates at the city/regency level in West Java are found in Bogor, Bekasi City, Bandung City, Bekasi, and Depok City. Conversely, Sumedang, Ciamis, Pangandaran, Majalengka, and Kuningan are the areas with the highest welfare. A distribution map shows that the regions bordering DKI Jakarta in the West area have higher poverty levels, while the eastern regions bordering Central Java tend to be more prosperous.

Poor households in West Java may not have access to credit (regional characteristics). However, it may have more than three household members living in one building (household characteristics), and the householder does not graduate from elementary school (individual characteristics). Based on existing data, no specific community factors have been identified as causes of poverty. Other predictor variables in the research do not show significant differences that can be considered poverty-inducing factors in West Java.

IV. CONCLUSIONS

The MCS-SMOTE model is tested on 18 scenarios of imbalanced and overlapping simulation data with multiclass response variables. The result shows good performance. As the level of imbalanced data decreases, the number of minority classes and overlapping data decreases, and the performance of the model improves. The proposed model produces values above 70% for all scenarios when evaluated using accuracy, F1 score, and G-Mean measures. Compared with five other models, the proposed model also provides the highest performance.

The level of imbalanced and overlapping data in poverty data is high, similar to the Extreme-2 simulation scenario with near distances. The MCS-SMOTE model performs well on poverty data, producing an accuracy value of 80.22%, an F1 score of 0.78, and a G-Mean of 0.21. The model's performance is satisfactory overall and separately, especially in classifying minority classes. Comparison with other models demonstrates the ability of the proposed model to solve existing problems. Evaluation of the model on training and test data reveals no significant differences, indicating that the model does not overfit the data.

The research focuses on analyzing multiclass data that has three variables. However, the empirical data are limited to a specific period in the West Java region. In future research, a new model can be developed that addresses the challenge of imbalance and overlap in data with more than three classes. It is important to note that the factors causing poverty in other provinces may differ from those discovered in West Java due to the regional aspect. Therefore, further research is recommended to sustainably analyze poverty factors in different regions.

ACKNOWLEDGMENT

REFERENCES

Aldania, A. N. A., Soleh, A. M., & Notodiputro, K. A. (2023). A comparative study of CatBoost and double random forest for multi-class classification. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, *7*(1), 129–137. https://doi.org/10.29207/resti.v7i1.4766

Badan Pusat Statistik Kabupaten Pesisir Selatan. (2023). *Garis kemiskinan (Rupiah), 2020-2022*. https://pesselkab.bps.go.id/indicator/23/96/1/garis-kemiskinan.html

Brereton, R. G. (2021). Contingency tables, confusion matrices, classifiers and quality of prediction. *Journal of Chemometrics*, *35*(11), 1–5. https://doi.org/10.1002/cem.3331

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

De Diego, I. M., Redondo, A. R., Fernández, R. R., Navarro, J., & Moguerza, J. M. (2022). General performance score for classification problems. *Applied Intelligence, 52*, 12049–12063. https://doi.org/10.1007/s10489-021-03041-7

Djamaluddin, S. (2017). How to lower the poverty?: Population control and increase of asset ownership. *Signifikan: Jurnal Ilmu Ekonomi*, *6*(2), 267–288. https://doi.org/10.15408/sjie.v6i2.5096

Esteves, V. M. S. (2020). *Techniques to deal with imbalanced data in multi-class problems: A review of existing methods*. Universidade Do Porto.

Fernández, A., García, S., Galar, M., Prati, R. C., Krawczyk, B., & Herrera, F. (2018). Imbalanced classification with multiple classes. In *Learning from imbalanced data sets* (pp. 197–226). Springer International Publishing. https://doi.org/10.1007/978-3-319-98074-4_8

Galar, M., Fernández, A., Barrenechea, E., Bustince, H., & Herrera, F. (2011). An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes. *Pattern Recognition*, *44*(8), 1761–1776. https://doi.org/10.1016/j.patcog.2011.01.017

Ishak, N. A., Ng, K. H., Tong, G. K., Kalid, S. N., & Khor, K. C. (2022). Mitigating unbalanced and overlapped classes in credit card fraud data with enhanced stacking classifiers system. *F1000Research*, *11*, 11–71.

Kalid, S. N., Ng, K. H., Tong, G. K., & Khor, K. C. (2020). A multiple classifiers system for anomaly detection in credit card data with unbalanced and overlapped classes. *IEEE Access*, *8*, 28210–28221. https://doi.org/10.1109/ACCESS.2020.2972009

Lango, M., & Stefanowski, J. (2022). What makes multi-class imbalanced problems difficult? An experimental study. *Expert Systems with Applications*, *199*. https://doi.org/10.1016/j.eswa.2022.116962

Meidianingsih, Q., & Meganingtyas, D. E. W. (2022). Analisis perbandingan performa metode ensemble dalam menangani imbalanced multi-class classification. *Jurnal Aplikasi Statistika dan Komputasi Statistik*, *14*(1), 13–21. https://doi.org/10.34123/jurnalasks.v14i2.335

Ongko, E., & Hartono. (2021). Hybrid approach redefinition-multi class with resampling and feature selection for multi-class imbalance with overlapping and noise. *Bulletin of Electrical Engineering and Informatics*, *10*(3), 1718–1728. https://doi.org/10.11591/eei.v10i3.3057

Pensasaran Percepatan Penghapusan Kemiskinan Ekstrem. (2022). Tanya/jawab layanan data P3KE. Kementerian Koordinator Bidang Pembangunan Manusia dan Kebudayaan. https://p3ke.kemenkopmk.go.id/tanyajawab/

Pradana, Z. H., Nafi'ah, H., & Rochmanto, R. A. (2022). Chatbot-based information service using RASA open-source framework in Prambanan Temple tourism object. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, *6*(4), 656–662. https://doi.org/10.29207/resti.v6i4.3913

Rosita, A. A., Kurnia, A., & Djuraidah, A. (2022). Evaluation of ensemble method for multiclass classification on unbalanced data. In *AIP Conference Proceedings* (Vol. 2662, No. 1). AIP Publishing. https://doi.org/10.1063/5.0108842

Sharma, N., Chakrabarti, A., Balas, V. E., & Martinovic, J. (Eds.). (2021). *Data management, analytics and innovation: Proceedings of ICDMAI 2020* (Vol. 2). Springer. https://doi.org/10.1007/978-981-15-5619-7

Tanha, J., Abdi, Y., Samadi, N., Razzaghi, N., & Asadpour, M. (2020). Boosting methods for multi-class imbalanced data classification: An experimental review. *Journal of Big Data*, *7*, 1–47. https://doi.org/10.1186/s40537-020-00349-y

Vuttipittayamongkol, P., Elyan, E., & Petrovski, A. (2021). On the class overlap problem in imbalanced data classification. *Knowledge-Based Systems*, *212*. https://doi.org/10.1016/j.knosys.2020.106631